

**Tobias Müller,
Olaf Krieger,
Klaus Lange,
Andreas Breuer,
Thomas Form**

**Neuronale Netzmengen für die Fehlerdiagnose in komplexen
Fahrzeugsystemen**

**Braunschweig : Institut für Regelungstechnik
Wolfsburg : Volkswagen AG**

Veröffentlicht: 25.07.2008

<http://www.digibib.tu-bs.de/?docid=00022899>

Auch erschienen in:
Diagnose in mechatronischen Fahrzeugsystemen, Expert Verlag,
2008, ISBN 978-3-8169-2821-8

Neuronale Netzmengen für die Fehlerdiagnose in komplexen Fahrzeugsystemen

Tobias Müller, Olaf Krieger, Klaus Lange, Andreas Breuer, Thomas Form

Abstract

Incremental insertion of complex mechatronic systems in the vehicles leads to a serious need of a parallel development of new methods and approaches for a safe-based diagnosis in the vehicles' factory. The idea behind this is to improve both the detection and the evaluation processes of faulty systems using new techniques for the diagnosis process. The adopted approach is to enable the diagnostic system to learn from the data of the past repairing cases based on neural networks, which involve artificial intelligence methodologies. These methodologies enable the diagnostic system to learn -from the coherence and independences involved in the past data- the needed know how, which can be further used for future repairing cases.

Kurzfassung

Die Komplexität heutiger Fahrzeuge ist in den letzten Jahren stark gestiegen. Dies sorgt für immer schwerer diagnostizierbare Fahrzeuge, an denen die heute eingesetzten Offboard-Diagnosesysteme an ihre Grenzen stoßen. Es sind neue innovative Diagnoseverfahren notwendig, um die Komplexität zukünftig handhaben zu können. Der Vortrag stellt ein neues heuristisches Diagnoseverfahren vor, welches mittels Mengen neuronaler Netze aus vergangenen Reparaturfällen lernt und das erlernte Wissen für künftige Reparaturfälle zur Unterstützung der Fehlersuche zur Verfügung stellt.

1. Einleitung

Moderne Fahrzeuge verfügen über eine Vielzahl an Steuergeräten und elektronischen Komponenten, die ein unkompliziertes, bequemes und sicheres Fahren gewährleisten. Dies hat zu einem deutlichen Anstieg der Komplexität geführt, die eine große Herausforderung für die in der Praxis eingesetzten Diagnosesysteme darstellt. Während bei Steuergeräten der ersten Generation jede Funktion genau einem Steuergerät zugeordnet war, werden Funktionen heute auf diverse Steuergeräte verteilt. Zusätzlich wird die Diagnose durch die Tatsache erschwert, dass ein Symptom nicht selten mehrere Ursachen hat, und eine Ursache oft zu mehreren Symptomen führt, vgl. [2], [3]. Moderne Diagnosesysteme müssen in der Lage sein, trotz dieser gestiegenen Komplexität die Fehlersuchzeiten zu verkürzen.

2. Hintergrund und Ausgangssituation

Die meisten heute eingesetzten Steuergeräte sind eigendiagnosefähig. Das bedeutet, sie besitzen Diagnoseroutinen innerhalb der Steuergerätesoftware, welche die Sensoren und Aktoren gegeneinander plausibilisieren. Unter anderem kommen hier modellbasierte Diagnoseverfahren zum Einsatz [5]. Wird vom System eine Störung erkannt, so wird ein Ereignisspeichereintrag (DTC¹) hinterlegt, der über eine Diagnoseschnittstelle aus dem Fahrzeug ausgelesen werden kann [4].

Im Gegensatz zur Eigendiagnose (On-Board-Diagnose) versteht man unter Off-Board-Diagnose-Systemen eine Klasse von Systemen, die in der Werkstatt an ein Fahrzeug angeschlossen werden und von außen auf die Systeme im Fahrzeug zugreifen. Sie haben die Möglichkeit, Ereignisspeichereinträge aller vorhandenen Fahrzeugsysteme auszulesen und dem Kundendienstmitarbeiter in lesbarer Form anzuzeigen.

Leistungsfähige Diagnosesysteme sollen dem KD-Mitarbeiter nicht nur alle Informationen aus den einzelnen Steuergeräten anzeigen, sondern ihn zusätzlich bei der Fehlersuche aktiv unterstützen. Ein Beispiel hierfür im Kundendienst von Volkswagen ist die Geführte Fehlersuche (GFS). Die GFS verwendet Fehlersuchprogramme auf Basis von Entscheidungsbäumen, welche von Experten erstellt, in verschiedene Sprachen übersetzt und weltweit an die Anwender verteilt werden. Hier wird für jeden Fehlercode, der in einem Fahrzeug auftreten kann, ein Fehlersuchprogramm erstellt. Eine besondere Schwierigkeit besteht hierbei in der Abdeckung aller Varianten eines Fahrzeuges, welche sich durch die zahlreichen Ausstattungsmöglichkeiten ergeben. Dies ist bei der Erstellung der Fehlersuchprogramme zu berücksichtigen. So äußert sich derselbe Fehler in den unterschiedlichen Fahrzeugvarianten jeweils in anderer Art und auch Prüf- und Reparaturanweisungen können stark variieren.

Bereits im Voraus alle möglichen Fehlerquellen zu erkennen und abzudecken ist eine weitere Schwachstelle solcher manuell erstellter Fehlersuchprogramme.

Die genannten Punkte führen in der Praxis zwangsläufig zu Kompromissen, da das Berücksichtigen aller möglichen Fehlerquellen sowie aller Fahrzeugvarianten nur mit erheblichem Aufwand beherrschbar ist. Insbesondere der bereits genannte Anstieg an Komplexität als auch die Tendenz zu verteilten Funktionen führen damit vermehrt zu reduzierten und unspezifischeren Fehlersuchprogrammen (z.B. „Suche nach Stromlaufplan“).

Ein wichtiger Schritt zur Verbesserung der Diagnosequalität ist die Möglichkeit die Vorgänge im Feld auszuwerten, wie z.B. durch die Einführung einer Datenbank für Felddaten. So könnten z.B., die für einen Reparaturfall relevanten Daten wie Ereignisspeichereinträge u.s.w., aufgezeichnet werden. Mit den heute verfügbaren Informationen werden bereits Auswertungen zur Verbesserung der Prüfabläufe durchgeführt. Dieser Vorgang der Auswertung, wie auch die anschließende Verbesserung der Fehlersuchprogramme wird jedoch manuell vorgenommen und lässt sich nicht mit vertretbarem Aufwand automatisieren. Daher kann über diese Vorgehensweise nur eine geringe Anzahl an Prüfprogrammen verbessert werden. An diesem Punkt setzt das in diesem Beitrag nachfolgend beschriebene erfahrungsbasierte Diagnosesystem an.

¹ DTC – engl. Diagnostic Trouble Code

3. Die Idee

Weltweit existieren mehrere zehntausend KFZ-Werkstätten mit insgesamt über hunderttausend Mitarbeitern. Aufgrund dieses enormen Potentials, ist der Austausch an Erfahrung zwischen den einzelnen Werkstätten eine Herausforderung.

Dies soll mit einem Beispiel veranschaulicht werden. Im Fall eines komplexen Reperaturproblems für dessen Lösung ein KD-Mitarbeiter A einiges an Fehlersuchzeit und auf Verdacht getauschte Teile aufgewendet hat, bedeutet dies für den KD-Mitarbeiter B einer anderen Werkstatt mit demselben Problem, denselben Aufwand zu investieren. KD-Mitarbeiter B profitiert folglich nicht von der Erfahrung des KD-Mitarbeiters A. Wäre es möglich, die einmal gemachten Erfahrungen immer zur Verfügung zu haben, bestände ein großes Einsparpotential.

Das Ziel sollte daher sein, aus Reparaturfällen zu lernen und das erlangte Wissen für zukünftige Reparaturfälle zu nutzen. Da auch hier analog zur GFS die Menge der Daten und das daraus erlangbare Wissen groß ist, kann sowohl das Lernen als auch die Anwendung nur dann von gewinnbringender Effizienz sein, wenn der gesamte Prozess automatisch durchführbar ist.

4. Problemanalyse

Für ein lernendes System ist die Art und die Qualität der vorhandenen Daten von entscheidender Bedeutung. Gelernt werden soll aus den Reparaturfällen in den Werkstätten. Auch wenn eine Reparatur in der Werkstatt ein iterativer Prozess ist, so kann der Informationsgehalt auf eine Symptombild – Reparaturmaßnahmenrelation komprimiert werden (vgl. Bild 1). In einem Symptombild werden alle während der Reparatur festgestellten Symptome zusammengefasst. Hierzu zählen nicht nur

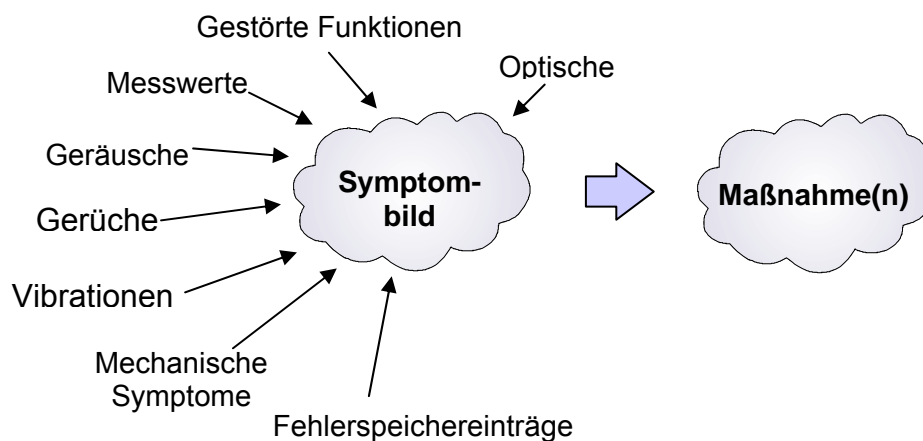


Bild 1: Abstraktion aus der zu lernenden Erfahrung.

technisch leicht und eindeutig erfassbare Symptome wie DTCs sondern auch subjektiv wahrgenommene Symptome welche in einer geeigneten Weise codiert wurden. Zu den Maßnahmen werden Aktionen des KD-Mitarbeiters, wie getauschte Teile oder Arbeitspositionen zusammengefasst. Auf diese Weise kann eine Reparatur als abgeschlossener Prozess mit einem Symptombild als Eingangsgröße und einer Reihe von Reparaturmaßnahmen als Ausgangsgröße betrachtet werden. Das dem Prozess zugrundeliegende System ist hierbei das jeweilige Fahrzeug an

sich. Nun werden keinesfalls immer absolut baugleiche Fahrzeuge repariert, vielmehr sind alle Fahrzeuge, selbst wenn sie von derselben Klasse sind, aufgrund der möglichen Varianten unterschiedlich. Als eine weitere Eingangsgröße des Reparaturprozesses sind folglich die fahrzeugspezifischen Eigenschaften als Kontext zu betrachten. Kontext und Symptombild werden hierbei zu einem Eingangsvektor zusammengefasst. Die resultierende Lernaufgabe besteht nun darin, aus den vorhandenen Daten ein Modell abzuleiten, das die beobachteten Eingangs- und Ausgangsgrößen bestmöglich abbildet.

Die Untersuchung verschiedener Technologien am IFR² hat ergeben, dass Künstliche Neuronale Netze (KNN) hierfür besonders gut geeignet sind. Auf eine Einführung in KNN wird an dieser Stelle verzichtet und auf [6] verwiesen. Eine wichtige Eigenschaft der KNN bei der zugrundeliegenden Aufgabe ist das Erkennen von Mustern in den vorhandenen Trainingsdaten (Patterns). Eine weitere wichtige Eigenschaft ist die Fähigkeit zu generalisieren, d.h. dass KNN ist in der Lage auch für Symptombilder, die nicht Teil der Trainingspatterns sind eine Aussage treffen zu können.

5. Realisierung

Ziel der Realisierung ist ein geschlossener Diagnoseprozess, der eine automatische Rückführung und Nutzung von Erfahrung ermöglicht. Ein prinzipielles Datenflussdiagramm ist in Bild 2 dargestellt. Im Werkstatt-Prozess sind alle Prozesse innerhalb einer Werkstatt zusammenzufassen. Eine Auswahl der entstehenden Daten, gemeint sind hier vor allem Reparaturfalldaten, ist in die Datenbank zu übertragen. Aus dem Datenbestand ist nun das Lernsystem zu speisen, welches aus der Vielzahl der Daten, wie vorhergehend beschrieben, Modelle erzeugt. Diese Modelle sind anschließend als Diagnosewissen ebenfalls in eine Datenbank zu übertragen, welche die Daten an die Werkstätten verteilt. Es handelt sich somit um einen geschlossenen Prozess, dessen Rückkopplung weitere Eigenschaften ausbildet. Hierdurch würde für bisher unbekannte Reparaturprobleme durch die Gesamtheit der KD-Mitarbeiter ein automatischer, iterativer Problemlösungsprozess entstehen. Dieser soll nachfolgend näher beschrieben werden.

Würde in einer Werkstatt ein neues bisher in den Datenbeständen unbekanntes Reparaturproblem auftreten, so besteht für dieses erste Auftreten im erlernten Modell kein Diagnosewissen. Eine zielgerichtete Diagnose würde nicht möglich sein. Der betreffende KD-Mitarbeiter müsste das Problem dann mit den ihm vorhandenen Mitteln lösen (z.B. Fehlersuchprogramme etc.). Weiterhin wird davon ausgegangen, dass in dieser ersten Problemlösung mehrere, auch z.T. nicht erfolgreiche Reparaturmaßnahmen erbracht werden, die anschließend in den Datenbestand einfließen. Im folgenden Schritt würde das Lernsystem erneut aus den Datenbeständen des Datawarehouse lernen und das Diagnosewissen optimieren. Würde nun dieses Problem in einer anderen Werkstatt erneut auftreten, so könnte das Diagnosesystem aufgrund der veränderten Modelle bereits eine zielgerichtete

² Institut für Regelungstechnik, TU Braunschweig

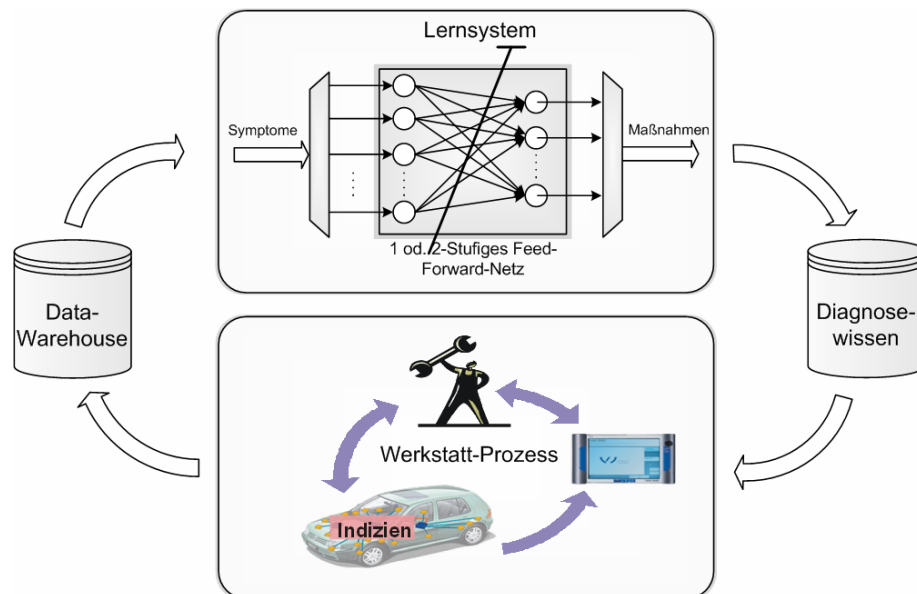


Bild 2: Datenflussdiagramm des Diagnose-Prozesses.

Diagnose stellen. Diese könnte aufgrund der Datenlage jedoch noch unscharf sein. Der betreffende KD-Mitarbeiter wird also im Regelfall eine Auswahl an den vorgeschlagenen Diagnosen treffen müssen. Wäre die „Reparaturlösung“ dabei Teil der Diagnosen, so würde die Menge der durchgeführten, nicht erfolgreichen Reparaturmaßnahmen im Mittel kleiner. Bereits nach ein paar wenigen Iterationen würde so automatisch die beste Reparaturlösung für das neue Problem gefunden. Nach dem gleichen Funktionsprinzip würden auf diese Weise nicht nur neue Problemlösungen gefunden sondern auch Veränderungen automatisch adaptiert werden.

6. Das Lernsystem

Eine entscheidende Rolle in dem beschriebenen Diagnoseprozess kommt dem Lernsystem zu. Eine nahe liegende Realisierung eines Lernsystems mit KNN ist in Bild 3 dargestellt. Es wird ein großes Feed-Forward-Netz ohne oder mit einem Hidden-Layer erzeugt. Die Zahl der Eingangsneuronen entspricht dabei der Größe des Eingangsvektors (Anzahl der möglichen Symptome und Kontextinformationen) und analog die Zahl der Ausgangsneuronen der Größe des Ausgangsvektors (Anzahl der möglichen Reparaturmaßnahmen).

Bei der Realisierung eines solchen Systems hat sich gezeigt, dass schnell ein zu großes Netz entsteht. Selbst wenn bei einem Fahrzeug der Mittelklasse ausschließlich DTCs als Symptome und getauschte Ersatzteile als Maßnahmen verwendet werden, entsteht ein Netz mit mind. 3.000 Eingangs- und 10.000 Ausgangsneuronen. Die daraus resultierende Anzahl der Verbindungen liegt je nach Anzahl der Hidden-Layer zwischen 30.000.000 und 50.000.000. Dies sind zu viele, um ein solches Netz effizient erzeugen und trainieren zu können. Hinzu kommt, dass eine Maßnahme in den meisten Fällen lediglich von ein paar wenigen Symptomen abhängt und somit der größte Teil der Verbindungen zu einem Ausgangsneuron ein Gewicht nahe Null besitzen. Genau diese große Zahl an nicht relevanten Verbindungen führt jedoch zu Einbußen in der Qualität des Netzes. Sinnvoller ist es, mehrere kleinere und besser handhabbare Netze zu erzeugen.

Die Unterteilung in mehrere kleinere Netze ist jedoch kein triviales Problem. Um die positiven Eigenschaften der Netze, wie Mustererkennung und Generalisierung beizubehalten, sollten in einem Netz möglichst diejenigen Symptome und Maßnahmen zusammengefasst werden, zwischen denen es einen Zusammenhang gibt.

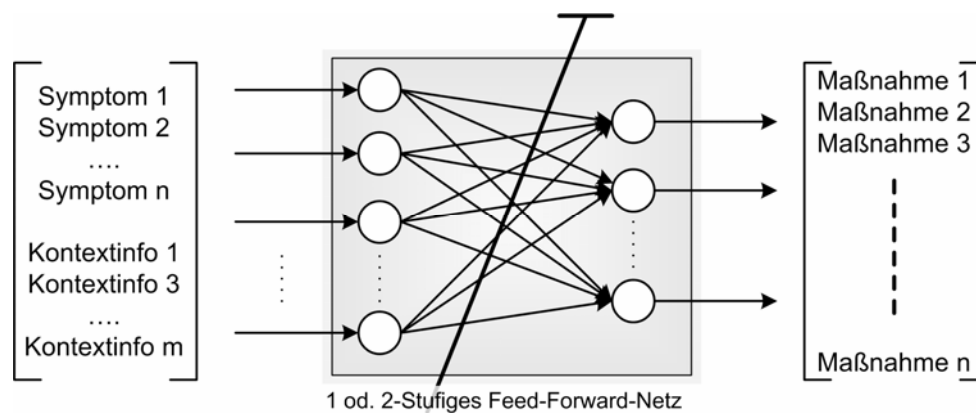


Bild 3: Lernsystem mit einem Neuronalen Netz.

Ein Maß für den Zusammenhang ist die aus der Statistik bekannte Signifikanz. Signifikanz bedeutet dabei, dass es mit einer gewissen Irrtumswahrscheinlichkeit p einen Zusammenhang zwischen Symptom und Maßnahme gibt. Ein p -Wert von 1 bedeutet eine 100%-tige Irrtumswahrscheinlichkeit für den Zusammenhang (es gibt also keinen Zusammenhang) und ein p -Wert von 0 bedeutet eine 0%-tige Irrtumswahrscheinlichkeit (es gibt einen eindeutigen Zusammenhang).

7. Clustern von Künstlichen Neuronalen Netzen

Für eine sinnvolle Unterteilung in kleinere Netze wurde bereits ein spezielles für die Anwendung zugeschnittenes Clusterverfahren entwickelt (vgl. [7] Abschnitt 7). Weitere Untersuchungen haben jedoch ergeben, dass dieses Verfahren speziell für größere Datenmengen einige Nachteile besitzt.

Das Verfahren neigt dazu mit zunehmender Datenzahl immer größere Cluster zu erzeugen. Dem kann zwar mit dem vorgestellten Kompensationsverfahren entgegengewirkt werden, jedoch verschlechtert sich dadurch auch die Qualität der Cluster zunehmend. Das Verfahren wurde konsequent weiterentwickelt um die beschriebenen Nachteile zu beseitigen und wird im Folgenden beschrieben.

Der erste Schritt ist analog zum bisherigen Verfahren das Finden von signifikanten Symptomen für jede Maßnahme. Mögliche Methoden für einen solchen Signifikanztest sind z.B. der χ^2 -Test (Chi-Quadrat-Test) nach (1) und der exakte Test nach Fisher [11]. Wird ein solcher Signifikanztest sukzessiv über den gesamten Datenbestand für alle Symptome und Maßnahmen durchgeführt, ergibt sich mit (3) die Signifikanzmatrix S . Aus der Signifikanzmatrix kann nun abgelesen werden, welche Symptome auf welche Maßnahme mit welcher Irrtumswahrscheinlichkeit (p -Wert) einen Einfluss hat. Der p -Wert wird dabei mit der Näherungsformel (2) aus der Prüfgröße χ^2 errechnet.

$$\chi^2 = \sum_{j=1}^u \sum_{k=1}^r \frac{(n_{jk} - n_{jk}^*)^2}{n_{jk}^*} \quad (1)$$

$$p = \frac{1}{2} \cdot 10^{\frac{\chi^2}{3,84}} \quad (2)$$

$$S = \begin{bmatrix} p_{m_1 s_1} & \dots & p_{m_1 s_i} \\ \vdots & \ddots & \vdots \\ p_{m_j s_1} & \dots & p_{m_j s_i} \end{bmatrix} \quad (3)$$

χ^2 = Prüfgröße für den χ^2 – Test

n_{jk} = Absolute Randhäufigkeit

n_{jk}^* = Erwartete Randhäufigkeit für H_0

u, r = Merkmalsdimension (hier 2)

In Schritt 2 wird die Signifikanzmatrix S in eine Distanzmatrix D überführt (vgl. (4)). Es wird dabei für jede Maßnahme und jedes Symptom berechnet welche Distanz d ein Symptom zu einer Maßnahme besitzt. Hierfür werden zunächst die einzelnen Distanzen (p -Werte) zwischen Maßnahmen und Symptomen als bipartitierter, ungerichteter, gewichteter Graph G nach (5) dargestellt, wobei sich für die Knotenmenge (6) eine disjunkte/echte Partitionierung in Maßnahmenknoten V_M und Symptomknoten V_S ergibt. Die Kanten E werden nach (7) jeweils mit den Kantengewichten $p_{s,m}$ aus der Signifikanzmatrix S für Kanten $e_{s,m} \in E$ definiert. Der kürzeste Weg zwischen Maßnahme m_i und Symptom s_j ist nun gegeben als Kantenmenge (8). Damit ergibt sich die Summe der Kantengewichte für den kürzesten Weg zwischen m_i und s_j mit (9). In Bild 4 ist dies beispielhaft für m_1 und s_8 ³ dargestellt. Die zugehörige Gleichung ist mit (10) definiert und zur Veranschaulichung nicht gemäß der Matrix S geordnet.

$$D = \begin{bmatrix} d_{m_1 s_1} & \dots & d_{m_1 s_i} \\ \vdots & \ddots & \vdots \\ d_{m_j s_1} & \dots & d_{m_j s_i} \end{bmatrix} \quad (4)$$

$$G = (V, E) \quad (5)$$

$$V = V_S \cup V_M \quad (6)$$

$$E \subseteq V_M \times V_S \quad (7)$$

$$K_{M_i, S_j} = \bigcup_{k=1..n} \{e_k \in E \mid e_1 = (m_i, s_1), e_2, \dots, e_{n-1}, e_n = (m_n, s_j)\} \quad (8)$$

$$d_{m_i s_j} = \sum_{e_{m,s} \in K_{M_i, S_j}} p_{m,s} \quad (9)$$

$$d_{m_1 s_8} = p_{m_1 s_3} + p_{s_3 m_2} + p_{m_2 s_7} + p_{s_7 m_3} + p_{m_3 s_8} \quad (10)$$

Die Notwendigkeit Wege über andere Maßnahmenknoten hinweg zu wählen, ergibt sich aus der Verteilung der Falldaten. Typische Auftretswahrscheinlichkeiten für ein beliebiges Symptom s_x unter der Annahme dass ein Fehlerfall F vorliegt, liegen im Bereich von $0,02\% \leq P(s_x | F) \leq 0,2\%$. Dadurch können sehr gut positive Zusammenhänge, jedoch keine Negativzusammenhänge festgestellt werden. Um dies zu

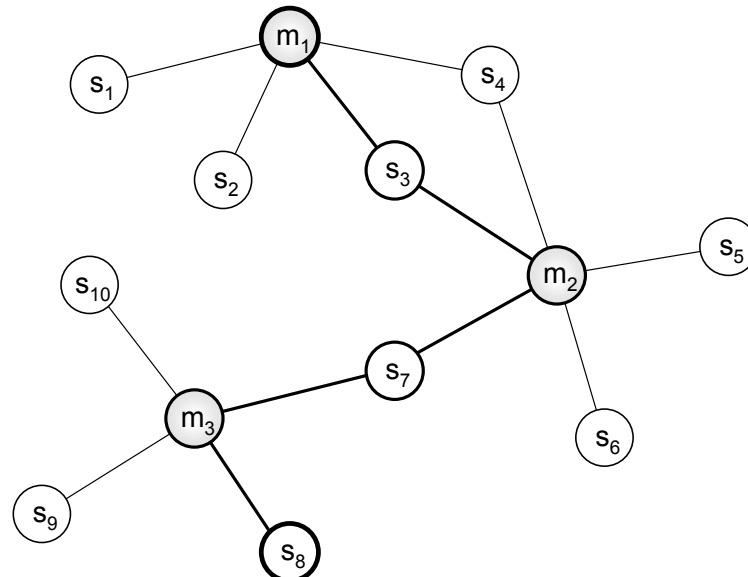


Bild 4: Beispielhaft dargestellte Distanz zwischen s_8 und m_1 ³.

zeigen wird zunächst die Definition des χ^2 -Test nach (1) für den hier konkret vorliegenden Fall mit 2 Variablen mit jeweils 2 Merkmalen (Symptom vorhanden (h_{x1}) / nicht vorhanden (h_{x2}); Maßnahme durchgeführt (h_{1x}) / nicht durchgeführt (h_{2x})) nach (11) überführt. Dieser Spezialfall wird auch χ^2 -Vierfeldertest genannt. Die χ^2 -Prüfgröße ist jedoch für einen anschaulichen Vergleich ungeeignet und wird darum mit der Beziehung (12) in den Phi-Koeffizient (13) überführt. Der Phi-Koeffizient stellt im Gegensatz zur χ^2 -Prüfgröße ein normiertes Zusammenhangsmaß (auch Korrelation genannt) mit dem Wertebereich von $-1 \leq \Phi \leq 1$ dar. Ein $\Phi > 0$ bedeutet dabei ein positiver Zusammenhang und ein $\Phi < 0$ ein negativer Zusammenhang. Jetzt wird ersichtlich, dass für ein Maximum bzw. ein Minimum der Korrelation die Felder der Hauptdiagonalen oder Nebendiagonalen von H gleich null sein müssen. Für einen positiven Zusammenhang kann dies leicht erfüllt werden, da hier die Felder h_{12} und h_{21} null sein müssen. Darüber hinaus bewirkt der durch die Falldaten bedingt große Wert von h_{22} , dass diese zwei Felder weniger ins Gewicht fallen. Für einen negativen Zusammenhang müssen jedoch die Felder h_{11} und h_{22} null oder nahe null sein. Für h_{22} ist dies nicht möglich. Eine Extremwertuntersuchung von (13) ergibt, dass die maximal negative Korrelation für die Bedingung $h_{12} = h_{21}$ erreicht wird. Damit kann das Minimum für Φ mit (14) bestimmt werden. Für die vorhergehend bereits erwähnten typischen Auftretswahrscheinlichkeiten $P(s_x | F)$ kann der Wertebereich von Φ_{\min} auf $10^{-4} \leq \Phi_{\min} \leq 10^{-3}$ bestimmt werden. Damit wird deutlich, warum Negativzusammenhänge nur schwer bzw. nicht erkannt werden können. Unter der Annahme, dass ein solcher Negativzusammenhang für eine andere Maßnahme

³ s_8 steht für Symptom 8, m_1 steht für Maßnahme 1. Die Nummerierung ist für die Darstellung frei gewählt.

einen Positivzusammenhang darstellt, ist die Wegbildung über andere Maßnahmen hinweg eine Lösung für dieses Problem.

$$\chi^2 = \frac{n \cdot (h_{11}h_{22} - h_{12}h_{21})^2}{(h_{11} + h_{12}) \cdot (h_{21} + h_{22}) \cdot (h_{11} + h_{21}) \cdot (h_{12} + h_{22})} \quad (11)$$

$$\chi^2 = n \cdot \Phi^2 \quad (12)$$

$$\Phi = \frac{h_{11}h_{22} - h_{12}h_{21}}{\sqrt{(h_{11} + h_{12}) \cdot (h_{21} + h_{22}) \cdot (h_{11} + h_{21}) \cdot (h_{12} + h_{22})}} \quad (13)$$

$$\Phi_{\min} = -\frac{h_{12}}{h_{12} + h_{22}} \quad \text{für } h_{12} = h_{21} \text{ und } h_{11} = 0 \quad (14)$$

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \text{Häufigkeitstabelle}$$

n = Fallzahl

Die erzeugte Distanzmatrix D , definiert in (4), dient nun als Grundlage für die Clustererzeugung für Schritt 3. Während das in [7] vorgestellte Verfahren nicht nur mehrere Symptome, sondern auch mehrere Maßnahmen in einem Cluster zusammenfasst, existiert hier immer nur eine Maßnahme pro Cluster. Es kann gezeigt werden, dass im speziellen hier vorliegenden Fall die Generalisierung eines neuronalen Netzes durch die Trennung in je ein Einzelnetz pro Ausgangsneuron nicht beeinträchtigt wird. Es existiert folglich keine Notwendigkeit mehrere Maßnahmen in einem Cluster zusammenzufassen. Darüber hinaus hat dies hier u.a. den Vorteil, dass die Einzelnetze übersichtlicher und robuster werden. Dies gilt gerade für die Verwendung von Hidden-Neuronen⁴. Im Fall von einer zu gering gewählten Zahl an Hidden-Neuronen in Kombination mit nicht zusammenhängenden Ausgangsneuronen, können ungewollte Randeefekte entstehen, die die Generalisierungsleistung stark herabsetzen.

Der Kern jedes Clusters C_{m_i} gehört somit zu genau einer Maßnahme m_i . Über eine zuvor definierte Threshold-Distanz d_{th} werden nun gemäß (15) all diejenigen Symptome dem Cluster C_{m_i} hinzugefügt, deren Distanz unterhalb der Threshold-Distanz d_{th} liegen. In Bild 5 wird dies Anschaulich dargestellt.

$$C_{m_i} = \bigcup_{s \in V_s} \{d_{m_i s} \leq d_{th}\} \quad (15)$$

C_{M_i} = Cluster für die Maßnahme M_i

⁴ Neuronen in einem Hidden-Layer werden auch Hidden-Neuronen bezeichnet.

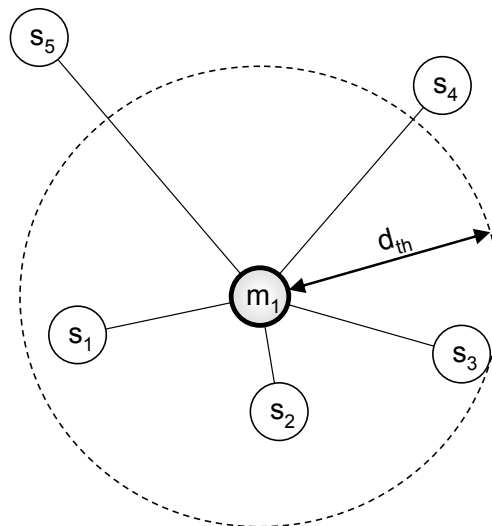


Bild 5: Beispielhaft dargestellte Distanzen und die Distanzgrenze für m_1 .

Bedingt durch die Definition des χ^2 -Tests (siehe (1) und [10]), sind die einzelnen p-Werte nach (2) und damit auch die Distanzwerte Fallzahlabhängig. Für die Bestimmung des Parameters d_{th} ist es jedoch wünschenswert, dass dieser nicht fallzahlabhängig bzw. normiert ist. Hierfür wird zunächst in (12) die Beziehung des χ^2 -Tests zum Phi-Koeffizienten betrachtet. Der Phi-Koeffizient stellt im Gegensatz zum χ^2 -Test ein normiertes Zusammenhangsmaß mit dem Wertebereich von $-1 \leq \Phi \leq 1$ dar. Ein Fallzahlunabhängiges Distanzmaß $d_{th,norm}$ wird folglich erreicht, indem die für die Berechnung notwendige Distanz d_{th} mit (1), (2) und (12) mit (16) definiert wird. Durch die Addition der einzelnen Distanzen im Graphen, kann $d_{th,norm}$ auch Werte > 1 annehmen.

$$d_{th} = \frac{1}{2} \cdot 10^{\frac{n \cdot d_{th,norm}^2}{3,84}} \quad (16)$$

n = Fallzahl

Nach dem Schritt der Clustererstellung wird anschließend in Schritt 4 für jedes Cluster ein eigenes KNN erstellt, wobei jedes Symptom durch ein Eingangsneuron und jede Maßnahme durch ein Ausgangsneuron repräsentiert wird. Die Vernetzung der Neuronen erfolgt dabei wie bereits in der vorhergehenden Realisierung Feed-Forward. Dies ist die klassische Konfiguration eines Pattern Associators [8]. Optional können noch Hidden-Layer eingefügt werden. Für die Aktivierungsfunktion hat sich bei den bisherigen Untersuchungen an exemplarischen Daten die sigmoide Funktion nach [9] als besonders vorteilhaft herausgestellt.

8. Zusammenfassung und Ergebnisse

Zum Entstehungszeitpunkt dieses Beitrags wurde bereits eine modulare Trainings- und Evaluierungssoftware erstellt, so dass bereits einige Versuche mit realen Daten durchgeführt werden konnten. Eine besonders schwierige Aufgabe war die Beschaffung der Daten. Weiterhin Zeitaufwendig war die Erfassung und Verknüpfung der Daten. Hinzu kommt, dass die Erzeugung der Daten in den Werkstätten an sich

nicht in der für ein Lernsystem nötigen Qualität stattfindet. Nur mit einigem Aufwand konnten die Daten verknüpft und ein Algorithmus entwickelt werden, der aus den Daten möglichst „qualitativ hochwertige“ Datensätze herausfiltert.

Ausgangsbasis für die Versuche war eine große Zahl realer Reparaturfälle. Als Symptome wurden DTCs und als Maßnahmen Ersatzteile, Arbeitspositionen und Komponentenbezeichnungen⁵ verwendet. Mit Hilfe der Trainingssoftware wurden neuronale Netze erzeugt, trainiert und anschließend mit den vorhandenen Daten validiert. Auf verschiedene Validierungsmethoden wurde bereits in [13] eingegangen. Des Weiteren wurde eine experimentelle Anwendungssoftware entwickelt, die es ermöglicht, die Diagnoseleistung der trainierten neuronalen Netze direkt am Fahrzeug zu erproben. Bild 6 zeigt einen Screenshot der Anwendungssoftware.

Für die Erprobung wurde unter anderem auch ein Praxistest durchgeführt, in dem gezielt Fehler in das Fahrzeug eingebaut und anschließend die durch das Diagnosesystem vorgeschlagenen Maßnahmen evaluiert wurden.

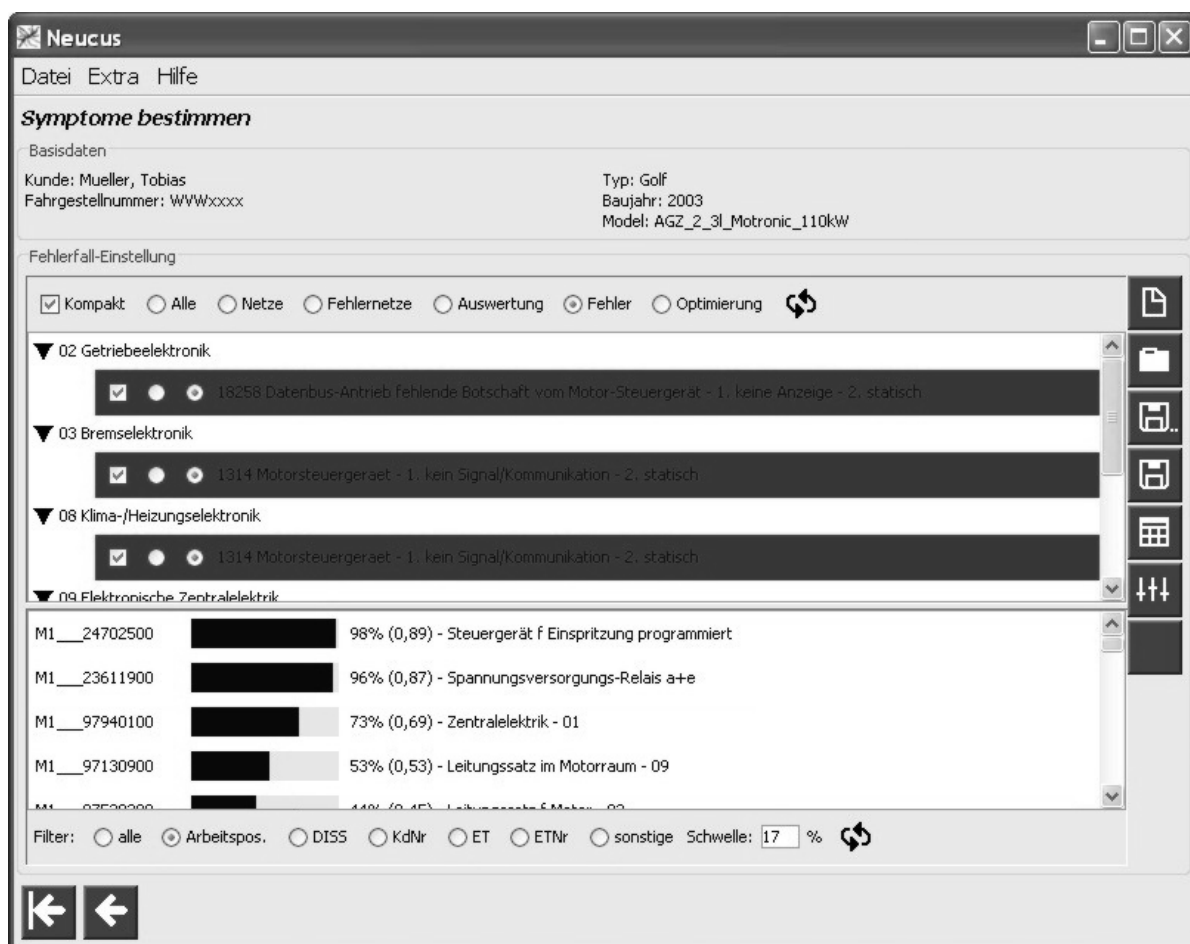


Bild 6: Screenshot der experimentellen Anwendungssoftware [12].

Die Tests zeigten, dass die neuronalen Netze nicht nur theoretisch sondern auch praktisch gute Ergebnisse liefern. Werden die verwendeten Daten betrachtet, zeigt sich, dass Neuronale Netze selbst in unscharfen Daten automatisch Muster und

⁵ Komponentenbezeichnungen = Ein für statistische Auswertungen eingeführter Katalog mit Reparaturmaßnahmen, die je einer Komponente z.B. Ölstandssensor zugeordnet sind.

Zusammenhänge erkennen, die auch praktisch für die Diagnose eingesetzt werden können. Die mittels KNN erzeugten Modelle eröffnen zusätzlich eine Bandbreite an neuen Möglichkeiten die Daten zu verstehen und z.B. für qualitätssteigernde Maßnahmen einzusetzen.

Darüber hinaus stellt ein erfahrungsbasiertes Diagnosesystem in der heutigen Diagnoselandschaft eine hervorragende Ergänzung zu den bisher eingesetzten Diagnosesystemen (z.B. GFS) dar. Es besitzt die entscheidenden Vorteile in der automatischen Generierung und Adaptierung von Diagnosewissen. Jedoch müssen für eine effiziente Integration in die bisherigen Prozesse noch einige Schritte getan werden. Letztlich ermöglicht ein solches Diagnosesystem jedoch auch in Zukunft immer komplexer werdende Fahrzeuge mit zahlreicher werdenden Varianten und verkürzten Entwicklungszyklen effizient instand zu setzen.

Literatur

- [1] Abuosba, Mohammad: Fahrzeug-Diagnose mit System. In: Sonderdruck aus Wissensmanagement 8/30 (2004)
- [2] Hack, Christian ; Drenkhahn, Georg ; Nickles, Jochen: Vernetzte Funktionen beherrschen - Diagnoseanforderung zur Bewältigung der Komplexität vernetzter Fahrzeugsysteme. In: Automotive, Messen und Testen (2006)
- [3] Rumpe, Bernhard; Schiffers, Jutta: Herausforderungen an die Diagnose - Integration der Diagnose in die Steuergeräteentwicklung. In: ZfAW. Zeitschrift für die gesamte Wertschöpfungskette Automobilwirtschaft (2006), S. 65–69
- [4] ISO 15031, Road vehicles - Communication between vehicle and external equipment for emissions-related diagnostics, ISO 2006
- [5] Isermann, R: Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren, Springer 2003
- [6] Callan, Robert: Neuronale Netze im Klartext. Pearson Studium Deutschland GmbH, 2006
- [7] T. Müller, K. Lange, A. Breuer, O. Krieger, T. Form: Automatische erfahrungsbasierte Diagnose aus Felddaten mit neuronalen Netzen, Elektronik im Kraftfahrzeug, Baden-Baden, 2007, ISBN 978-3-18-09200
- [8] Raul Rojas: Theorie der neuronalen Netze. Springer-Verlag, 1993
- [9] Christian Borgelt, Frank Klawonn, Rudolf Kruse: Neuronale Netze und Fuzzy-Systeme, 1994
- [10] Lothar Sachs: Angewandte Statistik: Anwendung statistischer Methoden, Springer-Verlag, 2004, ISBN 3540405550
- [11] Fisher, R. A.: Statistical Methods for research workers, Oliver and Boyd, 1954
- [12] Wittling, Daniel: Entwicklung und Optimierung von Anwendungs- und Validierungsmethoden für die Fehlerdiagnose mit neuronalen Netzen, Institut für Regelungstechnik, 2008, Braunschweig
- [13] T. Müller, O. Krieger, T. Form: Evaluierung von Offboard-Diagnosesystemen am Beispiel eines erfahrungsbasierten Diagnoseverfahrens, Simulation und Test, Berlin, 2008
- [14] O. Krieger, A. Breuer, T. Müller, T. Form, K. Lange: Wahrscheinlichkeitsbasierte Fahrzeugdiagnose auf Basis individuell generierter Prüfabläufe, Mechatronik 2007 - Innovative Produktentwicklung, 2007, ISBN 978-3-18-091971-3

Autoren

Dipl. Ing. Tobias Müller, Institut für Regelungstechnik, TU Braunschweig

Dipl. Ing. Olaf Krieger, Volkswagen AG, Wolfsburg

Dipl. Ing. Klaus Lange, Volkswagen AG, Wolfsburg

Dipl. Ing.(FH) Andreas Breuer, Volkswagen AG, Wolfsburg

Prof. Dr. Thomas Form, Institut für Regelungstechnik, TU Braunschweig